CS 232: Artificial Intelligence

Class Meetings:  M Th 8:30-9:45  
SCI 402

Instructor:  Carolyn Anderson  
carolyn.anderson@wellesley.edu  
SCI W421

Drop-in hours:  M Th 4-5:30pm  
Sci L-wing hallway

Tutor:  Funing Yang  

Tutor hours:  TBA  

Graders:  Shreya Kochar, Cynthia Wang

Announcements: We will use the Google Group TBA for course announcements. You are subscribed automatically. Please stay subscribed and check the list regularly.

Course Description
What is artificial intelligence (AI) and should humans fear it as one of "our biggest existential threats"? In this course, we will grapple with these difficult questions and investigate them in different ways. We will discuss the development of the field from the symbolic, knowledge-rich approaches of 20th century AI (e.g., rule-based systems), to statistical approaches that rely on increasingly large amounts of data, including an overview of contemporary deep learning techniques. We will explore how to apply these techniques in several AI application areas, including robotics, computer vision, and natural language processing, and consider ethical issues around AI in society. By the end of the semester, students should be able to answer the starting questions in-depth and with nuance.

Distributions: MM   Prerequisites(s): CS 230

Learning Goals

- Learn about a range of AI approaches, including rule-based/symbolic systems, statistical approaches, machine learning, and deep learning.
- Learn to recognize and choose appropriate techniques for a range of problems, including search, classification, and generation.
- Implement AI techniques in a variety of domains, including game-playing, robotics, computer vision, and natural language processing.
- Evaluate the performance of AI models with respect to both scientific validity and societal impact.
- Critically consider the ethical consequences of current and future AI technology.

Flexibility and Feedback
I welcome feedback throughout the course. You can either email me with your thoughts, or, if you prefer to be anonymous, you can submit them through the Anonymous Question Form.

Course Work
Expectations

- You are expected to come to class and to participate. This class involves in-class discussions and demonstrations, which are difficult to make up. However, you should not come to class if you are feeling sick (COVID or not); I’ll work individually with you to catch up after you are feeling better so that you don’t fall behind.

- Whenever you have questions about any course material, whether it is current material or from earlier in the course, please submit questions using the Anonymous Question Form. I strive to answer these questions within a day in the Q&A Document.

- You are expected to seek help when you have trouble with the course material, either from me or the course tutor. Our job is to help you succeed! Sometimes students are embarrassed to seek help, especially when they are very far behind. We are happy to help you with absolutely any material related to the course, no matter how long ago that material may have been covered.

Assignments

There will be 7 two-week homework assignments, typically composed of:

- short answer questions that ask you to reflect on the weekly readings
- short math questions to reinforce the theory behind the techniques we are studying
- programming exercises that you ask to implement AI algorithms or models
- evaluation questions that ask you to analyze the performance of a model that you have implemented

You should expect roughly 50% of each assignment to be Python programming.

Assignments will be due on Tuesdays at 10pm.

Many problem sets will also include one or more extra credit problems. These are fun but challenging. You should not attempt them until you have finished the required problems.

Final Project

In lieu of a final exam, there will be a final project. You will pick one of two topics: (1) cultural biases in large language models, or (2) the impact of language variation on language model performance. You will research one of these aspects and work to develop a probe task that can be applied to a neural network language model to investigate one aspect of the topic. Once you have developed your individual probe task, you will work with other group members to compile your findings into a research report. This report will be due at the end of finals period. Your group will present its findings in a short presentation on the last day of class.

Assignment Submission

We will use Google Drive for homework submissions. You will set up a Google Drive folder for your homework submissions and share it with me as part of Homework 1.
You will also fill out a Google Form for each assignment, which will ask you who you worked with (if anyone) and to estimate how much time you spend on the assignment.

Some homework problems will involve writing text, rather than code. You may intersperse these answers as comments in your Python program, or submit them as a separate PDF.

**Guest Lectures**

I have invited AI researchers to give research talks as part of this course. **Attending these talks is mandatory.** Some lectures may need to be held outside of normal class hours.

**Readings**

The main textbook for this course is *You Look Like A Thing and I Love You* by Janelle Shane. **Let me know if you have difficulty obtaining a copy.** Technical readings will come from the 3rd edition of *Artificial Intelligence: A Modern Approach* by Peter Norvig and Stuart Russell. **There is a PDF online.** Other occasional readings will be linked to on the course schedule.

**Collaboration**

Collaboration enriches the learning environment. I encourage you to talk with other students about the course material and to form study groups, with the following restrictions:

- **You may not share code with other students.** This includes code for assignments that you have already submitted.
- You may not look at code from other students.
- If you discuss a homework problem with another student, **please note which students** on your assignment when you submit it.

Some of these rules will be relaxed for the final assignment, which is a group project.

**Course Policies**

**Late Policy**

Late policies seek to balance two goals: providing flexibility for students, who often deal with unexpected life events and learning disruptions, and ensuring that students are able to progress in their learning when course content is cumulative.

The CS232 course content is highly cumulative. Also, some homework tasks are designed to help you explore a concept before I teach it; it is important that we are able to discuss your findings in subsequent classes. For this reason, the CS232 late policy will be relatively strict:

- Regular pssets are typically due at **10pm Tuesday.** For work submitted after this deadline, there will be **increasing penalties** based on how often you submit late work:
  - First time: no penalty
  - Second time: 10% penalty
  - Third time: 20% penalty
  - Fourth time: 40% penalty
Subsequent submissions: 50% penalty

- **If you cannot make a Tuesday deadline, email me as soon as possible.** We will work together to figure out a plan that supports both your learning and your well-being. **You do not need to reveal any personal details to justify your request.** Our plan may involve additional meetings outside of class to ensure that you can catch back up.

**Collaboration Policy and the Honor Code**

*I have adapted this wording from Lyn Turbak*

- **Collaboration:**
  - You may discuss high-level ideas or strategies with other students, but you must report who you worked with when you submit your assignment.
  - You must **not** communicate detailed algorithms, implementations, code, formulae, or other detailed solution steps.
  - Wait 30 minutes after discussions with other students before writing your solution. This helps you know if you actually understand the solution.
  - You may not share code from homework problems with other students at any point in the course (even after the assignment deadline).

- **Reference:**
  - You may consult course material from CS232 Spring 2022 and Python documentation.
  - You may consult external reference resources for general concepts and techniques, provided you cite them.
  - You must cite all collaboration, assistance, and reference (excluding course materials and Python documentation) that you used to prepare your submission.
  - You must **not** consult solutions to this or any similar assignment from the current or previous semesters of CS232, books, or online resources.

- **Code reuse and adaptation:**
  - You may reuse and adapt provided starter code in your solution.
  - You may reuse and adapt code from CS323 materials in this term.
  - You must **not** reuse or adapt any other code.

- **After the semester:**
  - Some students may want to include work from CS232 in portfolios. **Please do not post homework solutions publicly** (on GitHub, your website, etc.). You may show your code to potential employers or others who are not students.
  - Key point: **Do not make your code accessible or viewable by current or future CS232 students.**

**Grading**

- **Homework** 80%
- **Final project:**
  - Prep work 5%
  - Probe task 5%
  - Presentation 5%
  - Report 5%

This course will comply with the [Wellesley College grading policy](#).
Seeking Help

Lots of help is available in this course!

- If you go 20-30 minutes without making progress, please ask for help!
- You are encouraged to post questions to the [Anonymous Question Form](#). However, you may not post code. For code-specific questions, email me directly.
- Come to tutor hours! You don’t need a specific question to attend.
- Come to my drop-in hours! Again, you don’t need to bring a specific question. I’m also happy to talk about AI-related questions beyond the course material, as long as no one is waiting for help on homework. If the posted time or location doesn’t work for you, email me to schedule a Zoom meeting.
- You can request a one-on-one tutor from the PLTC.

Disabilities and Accommodations

Thanks to Ada Lerner for some of this wording.

My job is to help you learn, and part of that is making accommodations for any disabilities you might have. Please feel free to speak with me about concerns or suggestions about how I can make the course more accessible to you. I will keep the details of our conversations confidential. Though you are welcome to share any details that will be helpful, you are never required to share any private details of your life with me.

If you have a condition, either long-term or temporary, that would benefit from reasonable academic adjustments, it is strongly recommended that you contact Accessibility and Disability Resources (ADR) to get a letter outlining your accommodation needs, and submit it to me. You should request accommodations as early as possible in the semester, or before the semester begins, since some situations require significant time for review and accommodation design.

If you need immediate accommodations, please arrange to meet with me as soon as possible. If you are unsure but suspect you may have an undocumented need for accommodations, Disability Services can provide assistance including screening and referral for assessments.

Accessibility and Disability Resources can be reached at accessibility@wellesley.edu, at 781-283-2434, by scheduling an appointment online at their website [www.Wellesley.edu/adr](http://www.Wellesley.edu/adr), or by visiting their offices on the 3rd floor of Clapp Library, rooms 316 and 315.

Disclosures of Discrimination and Sexual Misconduct

Wellesley College considers diversity essential to educational excellence, and we are committed to being a community in which each member thrives. The College does not allow discrimination or harassment based on race, color, sex, gender identity or expression, sexual orientation, ethnic or national origin or ancestry, physical or mental disability, pregnancy or any other protected status under applicable local, state or federal law.
If you or someone you know experiences discrimination or harassment, support is available:

- **Confidential reporting:** Health Services (781.283.2810), the Stone Center Counseling Service (781.283.2839), and Religious and Spiritual Life (781.283.2685) are not required to report allegations of sexual misconduct to the College.

- **Non-confidential reporting:**
  - You can let me know. **As a faculty member, I am a mandatory reporter:** I am obligated to report allegations of sex-based discrimination to the Title IX Office.
  - You can report directly to the Nondiscrimination/Title IX Office (781.283.2451) to receive support, and to learn more about your options for reporting.
  - You can report to the Wellesley College Police (Emergency: 781.283.5555, Non-emergency: 781.283.2121) if you believe a crime has been committed.